

JC20 Rec'd PCT/PTO 2 6 SEP 2005

SEQUENCE LISTING

<110> JANSEN, KATHRIN U.
 5 SCHULTZ, LOREN D.
 NEEPER, MICHAEL P.
 MARKUS, HENRY Z.

<120> OPTIMIZED EXPRESSION OF HPV 31 L1 IN
 10 YEAST

<130> 21188P

<150> PCT/US2004/008677
 15 <151> 2004-03-19

<150> 60/457,172
 <151> 2003-03-23

20 <160> 8

<170> FASTSEQ FOR WINDOWS VERSION 4.0

<210> 1
 25 <211> 1515
 <212> DNA
 <213> HPV31 L1 WILD-TYPE

<400> 1

30 ATGTCTCTGT GCGGCCTAG CGAGGCTACT GTCTACTTAC CACCTGTCCC AGTGTCTAAA 60
 GTTGTAAGCA CGGATGAATA TGTAACACGA ACCAACATAT ATTATCACGC AGGCAGTGCT 120
 AGGCTGCTTA CAGTAGGCCA TCCATATTAT TCCATACCTA AATCTGACAA TCCTAAAAAA 180
 ATAGTTGTAC CAAAGGTGTC AGGATTACAA TATAGGGTAT TTAGGGTTTCG TTTACCAGAT 240
 CCAAACAAAT TTGGATTTCC TGATACATCT TTTTATAATC CTGAAACTCA ACGCTTAGTT 300
 35 TGGGCCTGTG TTGGTTTAGA GGTAGGTCGC GGGCAGCCAT TAGGTGTAGG TATTAGTGGT 360
 CATCCATTAT TAAATAAATT TGATGACACT GAAAACTCTA ATAGATATGC CGGTGGTCCT 420
 GGCAC TGATA ATAGGGAATG TATATCAATG GATTATAAAC AAACACAACT GTGTTTACTT 480
 GGTTGCAAAC CACCTATTGG AGAGCATTGG GGTAAAGGTA GTCCTTGTAG TAACAATGCT 540

```

ATTACCCCTG GTGATTGTCC TCCATTAGAA TTAAAAAATT CAGTTATACA AGATGGGGAT 600
ATGGTTGATA CAGGCTTTGG AGCTATGGAT TTTACTGCTT TACAAGACAC TAAAAGTAAT 660
GTCCTTTTGG ACATTTGTAA TTCTATTTGT AAATATCCAG ATTATCTTAA AATGGTTGCT 720
GAGCCATATG GCGATACATT ATTTTTTTTAT TTACGTAGGG AACAAATGTT TGTAAGGCAT 780
5 TTTTTTAATA GATCAGGCAC GGTGGTGAA TCGGTCCCTA CTGACTTATA TATTAAAGGC 840
TCCGGTTCAA CAGCTACTTT AGCTAACAGT ACATACTTTC CTACACCTAG CGGCTCCATG 900
GTTACTTCAG ATGCACAAAT TTTTAATAAA CCATATTGGA TGCAACGTGC TCAGGGACAC 960
AATAATGGTA TTTGTTGGGG CAATCAGTTA TTTGTTACTG TGGTAGATAC CACACGTAGT 1020
ACCAATATGT CTGTTTGTGC TGCAATTGCA AACAGTGATA CTACATTTAA AAGTAGTAAT 1080
10 TTTAAAGAGT ATTTAAGACA TGGTGAGGAA TTTGATTTAC AATTTATATT TCAGTTATGC 1140
AAAATAACAT TATCTGCAGA CATAATGACA TATATTCACA GTATGAATCC TGCTATTTTG 1200
GAAGATTGGA ATTTTGGATT GACCACACCT CCCTCAGGTT CTTTGGAGGA TACCTATAGG 1260
TTTGTAACCT CACAGGCCAT TACATGTCAA AAAAGTGCCC CCCAAAAGCC CAAGGAAGAT 1320
CCATTTAAAG ATTATGTATT TTGGGAGGTT AATTTAAAAG AAAAGTTTTTC TGCAGATTTA 1380
15 GATCAGTTTC CACTGGGTCG CAAATTTTTTA TTACAGGCAG GATATAGGGC ACGTCCTAAA 1440
TTTAAAGCAG GTAAACGTAG TGCACCCTCA GCATCTACCA CTACACCAGC AAAACGTAAA 1500
AAACTAAAA AGTAA 1515

```

<210> 2

20 <211> 1515

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

25 <223> 31 PARTIAL REBUILD

<400> 2

```

ATGTCTCTGT GGCGGCCTAG CGAGGCTACT GTCTACTTAC CACCTGTCCC AGTGTCTAAA 60
GTTGTAAGCA CGGATGAATA TGTAACACGA ACCAACATAT ATTATCACGC AGGCAGTGCT 120
30 AGGCTGCTTA CAGTAGGCCA TCCATATTAT TCCATACCTA AATCTGACAA TCCTAAAAAA 180
ATAGTTGTAC CAAAGGTGTC AGGATTACAA TATAGGGTAT TTAGGGTTTCG TTTACCAGAT 240
CCAAACAAAT TTGGATTTC TGTACATCT TTTTATAATC CTGAAACTCA ACGCTTAGTT 300
TGGGCCTGTG TTGGTTTAGA GGTAGGTCGC GGGCAGCCAT TAGGTGTAGG TATTAGTGGT 360
CATCCATTAT TAAATAAATT TGATGACACT GAAACTCTA ATAGATATGC CGGTGGTCCT 420
35 GGCCTGATA ATAGGGAATG TATATCAATG GATTATAAAC AAACACAACT GTGTTTACTT 480
GGTTGCAAC CACCTATTGG AGAGCATTGG GGTAAAGGTA GTCCTTGTAG TAACAATGCT 540
ATTACCCCTG GTGATTGTCC TCCATTAGAA TTAAAAAATT CAGTTATACA AGATGGGGAT 600
ATGGTTGATA CAGGCTTTGG AGCTATGGAT TTTACTGCTT TACAAGACAC TAAAAGTAAT 660

```

```

G TTCCTTTGG ACATTTGTAA TTCTATTTGT AAATATCCAG ATTATCTTAA AATGGTTGCT 720
GAGCCATACG GCGACACCTT GTTCTTCTAT TTGCGTAGAG AACAGATGTT CGTAAGGCAC 780
TTCTTCAACA GATCCGGCAC CGTAGGTGAA TCTGTCCCAA CCGACCTGTA CATCAAGGGC 840
TCCGGTTCCA CCGCTACCCT GGCTAACTCC ACCTACTTCC CAACTCCATC TGGCTCCATG 900
5  GTCACCTCCG ACGCTCAGAT CTTCAACAAG CCATACTGGA TGCAGCGTGC ACAGGGTCAC 960
AACACGGTA TCTGTTGGGG TAACCAGCTG TTCGTGACTG TGGTCGATAC CACGCGTTCT 1020
ACCAACATGT CTGTCTGTGC TGCAATCGCT AACTCTGACA CTACCTTCAA GTCCTCTAAC 1080
TTCAAGGAGT ACCTGAGACA TGGTGAGGAA TTCGATCTGC AATTCATCTT CCAGTTGTGC 1140
AAGATCACCC TGTCTGCTGA CATCATGACC TACATCCACA GTATGAACCC TGCCATCCTG 1200
10 GAGGACTGGA ACTTCGGTCT GACCACTCCA CCTTCCGGTT CTTTGGAGGA TACCTATAGG 1260
TTTGTAACCT CACAGGCCAT TACATGTCAA AAAAGTGCCC CCCAAAAGCC CAAGGAAGAT 1320
CCATTTAAAG ATTATGTATT TTGGGAGGTT AATTTAAAAG AAAAGTTTTT TGCAGATTTA 1380
GATCAGTTTC CACTGGGTCG CAAATTTTTA TTACAGGCAG GATATAGGGC ACGTCCTAAA 1440
TTTAAAGCAG GTAAACGTAG TGCACCCTCA GCATCTACCA CTACACCAGC AAAACGTAAA 1500
15 AAAACTAAAA AGTAA 1515

```

<210> 3

<211> 1515

<212> DNA

20 <213> ARTIFICIAL SEQUENCE

<220>

<223> 31 TOTAL REBUILD

25 <400> 3

```

ATGTCTTTGT GGAGACCATC TGAAGCTACC GTCTACTTGC CACCAGTCCC AGTCTCTAAG 60
GTCGTCTCTA CCGACGAATA CGTCACCAGA ACCAACATCT ACTACCACGC TGGTTCTGCT 120
AGATTGTTGA CCGTCGGTCA CCCATACTAC TCTATCCCAA AGTCTGACAA CCCAAAGAAG 180
ATCGTCGTCC CAAAGGTCTC TGGTTTGCAA TACAGAGTCT TCAGAGTCAG ATTGCCAGAC 240
30 CCAAACAAGT TCGGTTTCCC AGACACCTCT TTCTACAACC CAGAAACCCA AAGATTGGTC 300
TGGGCTTGTTG TCGGTTTGGA AGTCGGTAGA GGTCAACCAT TGGGTGTCGG TATCTCTGGT 360
CACCCATTGT TGAACAAGTT CGACGACACC GAAACTCTA ACAGATACGC TGGTGGTCCA 420
GGTACCGACA ACAGAGAATG TATCTCTATG GACTACAAGC AAACCAATT GTGTTTGTTG 480
GGTTGTAAGC CACCAATCGG TGAACACTGG GGTAAGGGTT CTCCATGTTT TAACAACGCT 540
35 ATCACCCAG GTGACTGTCC ACCATTGGAA TTGAAGAACT CTGTCATCCA AGACGGTGAC 600
ATGGTCGACA CCGGTTTCGG TGCTATGGAC TTCACCGCTT TGCAAGACAC CAAGTCTAAC 660
GTCCCATTGG ACATCTGTAA CTCTATCTGT AAGTACCCAG ACTACTTGAA GATGGTCGCT 720
GAACCATACG GCGACACCTT GTTCTTCTAC TTGCGTAGAG AACAGATGTT CGTAAGGCAC 780

```

```

TTCTTCAACA GATCCGGCAC CGTAGGTGAA TCTGTCCCAA CCGACCTGTA CATCAAGGGC 840
TCCGGTTCCA CCGCTACCCT GGCTAACTCC ACCTACTTCC CAACTCCATC TGGCTCCATG 900
GTCACCTCCG ACGCTCAGAT CTTCAACAAG CCATACTGGA TGCAGCGTGC ACAGGGTCAC 960
AACAACGGTA TCTGTTGGGG TAACCAGCTG TTCGTGACTG TGGTCGATAC CACGCGTTCT 1020
5 ACCAACATGT CTGTCTGTGC TGCAATCGCT AACTCTGACA CTACCTTCAA GTCCTCTAAC 1080
TTCAAGGAGT ACCTGAGACA TGGTGAGGAA TTCGATCTGC AATTCATCTT CCAGTTGTGC 1140
AAGATCACCC TGTCTGCTGA CATCATGACC TACATCCACA GTATGAACCC TGCCATCCTG 1200
GAGGACTGGA ACTTCGGTCT GACCACTCCA CCTTCCGGTT CTTTGAAGA CACCTACAGA 1260
TTCGTACCT CTCAAGCTAT CACCTGTCAA AAGTCTGCTC CACAAAAGCC AAAGGAAGAC 1320
10 CCATTCAAGG ACTACGTCTT CTGGGAAGTC AACTTGAAGG AAAAGTTCTC TGCTGACTTG 1380
GACCAATTCC CATTGGGTAG AAAGTTCTTG TTGCAAGCTG GTTACAGAGC TAGACCAAAG 1440
TTCAAGGCTG GTAAGAGATC TGCTCCATCT GCTTCTACCA CCACCCAGC TAAGAGAAAG 1500
AAGACCAAGA AGTAA 1515

```

```

15 <210> 4
    <211> 504
    <212> PRT
    <213> ARTIFICIAL SEQUENCE

```

```

20 <220>
    <223> HPV 31 L1

```

```

<400> 4
MET SER LEU TRP ARG PRO SER GLU ALA THR VAL TYR LEU PRO PRO VAL
25 1 5 10 15
PRO VAL SER LYS VAL VAL SER THR ASP GLU TYR VAL THR ARG THR ASN
    20 25 30
ILE TYR TYR HIS ALA GLY SER ALA ARG LEU LEU THR VAL GLY HIS PRO
    35 40 45
30 TYR TYR SER ILE PRO LYS SER ASP ASN PRO LYS LYS ILE VAL VAL PRO
    50 55 60
LYS VAL SER GLY LEU GLN TYR ARG VAL PHE ARG VAL ARG LEU PRO ASP
65 70 75 80
PRO ASN LYS PHE GLY PHE PRO ASP THR SER PHE TYR ASN PRO GLU THR
35 85 90 95
GLN ARG LEU VAL TRP ALA CYS VAL GLY LEU GLU VAL GLY ARG GLY GLN
    100 105 110
PRO LEU GLY VAL GLY ILE SER GLY HIS PRO LEU LEU ASN LYS PHE ASP

```

		115		120		125											
		ASP	THR	GLU	ASN	SER	ASN	ARG	TYR	ALA	GLY	GLY	PRO	GLY	THR	ASP	ASN
		130						135					140				
		ARG	GLU	CYS	ILE	SER	MET	ASP	TYR	LYS	GLN	THR	GLN	LEU	CYS	LEU	LEU
5		145					150					155				160	
		GLY	CYS	LYS	PRO	PRO	ILE	GLY	GLU	HIS	TRP	GLY	LYS	GLY	SER	PRO	CYS
						165					170				175		
		SER	ASN	ASN	ALA	ILE	THR	PRO	GLY	ASP	CYS	PRO	PRO	LEU	GLU	LEU	LYS
					180					185				190			
10		ASN	SER	VAL	ILE	GLN	ASP	GLY	ASP	MET	VAL	ASP	THR	GLY	PHE	GLY	ALA
				195					200					205			
		MET	ASP	PHE	THR	ALA	LEU	GLN	ASP	THR	LYS	SER	ASN	VAL	PRO	LEU	ASP
				210					215					220			
		ILE	CYS	ASN	SER	ILE	CYS	LYS	TYR	PRO	ASP	TYR	LEU	LYS	MET	VAL	ALA
15		225					230					235				240	
		GLU	PRO	TYR	GLY	ASP	THR	LEU	PHE	PHE	TYR	LEU	ARG	ARG	GLU	GLN	MET
						245					250				255		
		PHE	VAL	ARG	HIS	PHE	PHE	ASN	ARG	SER	GLY	THR	VAL	GLY	GLU	SER	VAL
					260				265					270			
20		PRO	THR	ASP	LEU	TYR	ILE	LYS	GLY	SER	GLY	SER	THR	ALA	THR	LEU	ALA
				275					280					285			
		ASN	SER	THR	TYR	PHE	PRO	THR	PRO	SER	GLY	SER	MET	VAL	THR	SER	ASP
				290					295				300				
		ALA	GLN	ILE	PHE	ASN	LYS	PRO	TYR	TRP	MET	GLN	ARG	ALA	GLN	GLY	HIS
25		305					310					315				320	
		ASN	ASN	GLY	ILE	CYS	TRP	GLY	ASN	GLN	LEU	PHE	VAL	THR	VAL	VAL	ASP
						325					330				335		
		THR	THR	ARG	SER	THR	ASN	MET	SER	VAL	CYS	ALA	ALA	ILE	ALA	ASN	SER
					340				345					350			
30		ASP	THR	THR	PHE	LYS	SER	SER	ASN	PHE	LYS	GLU	TYR	LEU	ARG	HIS	GLY
				355					360					365			
		GLU	GLU	PHE	ASP	LEU	GLN	PHE	ILE	PHE	GLN	LEU	CYS	LYS	ILE	THR	LEU
				370				375					380				
		SER	ALA	ASP	ILE	MET	THR	TYR	ILE	HIS	SER	MET	ASN	PRO	ALA	ILE	LEU
35		385					390					395				400	
		GLU	ASP	TRP	ASN	PHE	GLY	LEU	THR	THR	PRO	PRO	SER	GLY	SER	LEU	GLU
						405					410				415		
		ASP	THR	TYR	ARG	PHE	VAL	THR	SER	GLN	ALA	ILE	THR	CYS	GLN	LYS	SER

420 425 430
 ALA PRO GLN LYS PRO LYS GLU ASP PRO PHE LYS ASP TYR VAL PHE TRP
 435 440 445
 GLU VAL ASN LEU LYS GLU LYS PHE SER ALA ASP LEU ASP GLN PHE PRO
 5 450 455 460
 LEU GLY ARG LYS PHE LEU LEU GLN ALA GLY TYR ARG ALA ARG PRO LYS
 465 470 475 480
 PHE LYS ALA GLY LYS ARG SER ALA PRO SER ALA SER THR THR THR PRO
 485 490 495
 10 ALA LYS ARG LYS LYS THR LYS LYS
 500

<210> 5
 15 <211> 34
 <212> DNA
 <213> ARTIFICIAL SEQUENCE

<220>
 20 <223> PCR PRIMER

<400> 5
 CGTCGACGTA AACGTGTATC ATATTTTTTTT ACAG

34

25 <210> 6
 <211> 25
 <212> DNA
 <213> ARTIFICIAL SEQUENCE

30 <220>
 <223> PCR PRIMER

<400> 6
 CAGACACATG TATTACATAC ACAAC

25

35 <210> 7
 <211> 41
 <212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> PCR PRIMER

5

<400> 7

CTCAGATCTC ACAAACAAA ATGTCTCTGT GCGGCCTAG C

41

<210> 8

10

<211> 38

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

15

<223> PCR PRIMER

<400> 8

GACAGATCTT ACTTTT TAGT TTTT TACGT TTTGCTGG

38